

DOI: 10.19650/j.cnki.cjsi.J1905607

基于激光点云与图像信息融合的交通环境车辆检测^{*}

郑少武¹, 李巍华¹, 胡坚耀²

(1. 华南理工大学机械与汽车工程学院 广州 510641; 2. 工业和信息化部电子第五研究所 广州 510610)

摘要:随着无人驾驶汽车的快速发展,仅依靠单一传感器的环境感知已经无法满足车辆在复杂交通场景下的目标检测需求。融合多种传感器数据已成为无人驾驶汽车的主流感知方案。提出一种基于激光点云与图像信息融合的交通环境车辆检测方法。首先,利用深度学习方法对激光雷达和摄像头传感器所采集的数据分别进行目标检测;其次,利用匈牙利算法对两种目标检测结果进行实时目标跟踪,进而对两种传感器检测及跟踪结果的特征进行最优匹配;最后,将已匹配及未匹配的目标进行择优输出,作为最终感知结果。利用公开数据集 KITTI 的部分交通环境跟踪序列及实际道路测试来验证所提方法,结果表明,所提融合方法实时车辆检测精度提升 12% 以上、误检数减少 50% 以上。

关键词:激光点云;信息融合;环境感知;车辆检测

中图分类号: TP391.4 TH744 **文献标识码:** A **国家标准学科分类代码:** 520.20

Vehicle detection in the traffic environment based on the fusion of laser point cloud and image information

Zheng Shaowu¹, Li Weihua¹, Hu Jianyao²

(1. School of Mechanical & Automotive Engineering, South China University of Technology, Guangzhou 510641, China;

2. China CEPREI Laboratory, Guangzhou 510610, China)

Abstract: With the rapid development of driverless cars, the environment perception solution relying on single sensor cannot meet the demands of vehicles target detection in complex traffic scenarios. Fusion of multiple sensors has become a mainstream perception solution for driverless vehicles. In this study, a vehicle detection method in traffic environment based on the fusion laser point cloud and image information is proposed. Firstly, the deep learning method is used to detect the object data collected by the lidar and the camera sensor. Secondly, the Hungarian algorithm is utilized to track the target detection results in real-time. Then, the characteristics of the detection and tracking results from the two sensors are optimally matched with each other. Finally, the matched and unmatched targets are picked and outputted as the final perception results. The proposed algorithm is evaluated in some traffic environment tracking sequence of the public dataset KITTI and real road testing. Experimental results show that the real-time vehicle detection accuracy of the proposed fusion method increases more than 12% and the number of false detection decreases more than 50%.

Keywords: laser point cloud; information fusion; environment perception; vehicle detection

0 引言

智能化已成为汽车产业的一个重要发展趋势,越来越多的智能辅助驾驶功能正在逐渐改变人们的出行体验。随着智能化程度的不断提升,智能驾驶正朝着无人驾驶的方向快速发展。环境感知系统需要为无人驾驶汽

车提供准确、可靠的目标类别和位置信息,为无人驾驶路径规划及车辆控制提供可靠的决策信息^[1]。

车辆检测融合算法目前主要包括 2D 图像目标检测、3D 点云目标检测、激光与视觉融合的目标检测等。其中图像目标检测网络主要分为两大类:一种是基于候选框提取的目标检测算法,如区域-卷积神经网络(region-convolutional neural network, R-CNN)^[2]、快速区域卷积神

收稿日期:2019-09-15 Received Date:2019-09-15

^{*} 基金项目:广东省重点领域研发计划(2019B090912001,2019B090912002)项目资助

神经网络(Fast R-CNN)^[3]、Faster R-CNN^[4]等算法,将目标检测问题转换为分类问题,根据深度学习在大数据集上优异的分类性能提高检测精度;另一种是基于回归的目标检测算法,如SSD(single shot multibox detector)^[5]算法和YOLO(you only look once)系列^[6-7]算法,将目标检测中分类和定位问题融合到一个神经网络中处理,通过最小化网络损失函数实现对分类和定位的统一训练。根据现有研究的对比,YOLOv3算法在检测精度和速度上相比其他网络都具有较明显的优势,在COCO公开数据集上能够检测80种类别,可满足车辆检测的实时性要求^[7]。因此本文选用YOLOv3作为2D图像目标检测的基础框架。

点云目标检测网络可分为3种类型:1)是对原始点云转换为体素格式后输入网络进行检测;2)是将点云转换成2D图像后输入网络进行检测;3)是直接对原始点云进行特征学习后实现目标检测。Part-A² Net网络^[8]先对原始点云生成3D建议框,再对每个框内的点云做聚合并分组打分,输出点云的分类信息。VoxelNet网络^[9]对原始点云进行体素化,对体素特征进行学习后输出3D目标检测框。PointPillars网络^[10]首先将点云转换为稀疏的伪图像,再利用2D卷积网络对伪图像进行检测并回归成3D检测框。PointRCNN网络^[11]根据预测到的点云标签,对原始点云自下而上生成3D检测框,单帧检测速度为0.1 s,与激光雷达数据采集频率10 Hz一致,检测精度和检测速度相比其他网络有明显优势,在KITTI目标检测数据集上对车辆检测性能优秀。因此本文选择PointRCNN进行点云目标检测。

激光点云与图像信息融合算法在目标检测上可分为两种类型,一种是分别对激光点云和图像进行特征提取后,输入到同一神经网络中进行目标分类和预测,称为前融合方案。如Liang等^[12]提出了一种基于二维和三维目标检测、地面估计和深度补全的多任务多传感器检测模型;对图像进行特征提取后融合到点云鸟瞰图的特征图上进行3D目标检测^[13];MV3D(multi-

view 3D object detection network)网络^[14]将点云的前视图与鸟瞰图及图像同时作为网络输入,从鸟瞰图生成目标的提议框,并投影到3个视图进行特征提取,通过对特征进行融合来回归出3D检测框;聚合视图对象检测(aggregate view object detection, AVOD)网络^[15]分别对图像和点云鸟瞰图进行特征提取后,对特征进行融合,输出3D检测框。另一种是对激光点云和图像分别进行目标检测后,将输出的检测结果进行融合^[16-17],称为后融合方案。如Wu等^[17]分别利用点云和图像进行行人检测,并对检测结果进行匹配后输出最终融合的行人检测结果。

上述目标检测算法虽然取得了一定成果,但仍存在以下局限性:1)单一传感器难以应对复杂环境感知信息,相机无法获取精确的深度信息,雷达无法获取颜色、纹理信息,且检测精度仍有待进一步提升;2)从目标检测开源数据集KITTI排行榜数据可看出,现有大部分前融合方法网络本身较为复杂,难以实现实时性,实时应用效果不佳。3)后融合方法对不同的数据分别进行目标检测后再进行融合,融合方法相对灵活,但现有针对后融合方法的研究较少,应用较为局限,极少能够应用在交通环境车辆检测上。

针对以上问题,在环境感知系统中保证实时性的前提下,为了提高目标检测精度、降低误检率,本文提出一种通用的基于激光点云与图像信息融合的交通环境车辆检测方法,选取性能优异的目标检测算法作为基础网络,将目标检测与目标跟踪相结合,充分利用点云空间坐标与图像像素坐标的相互转换关系,实现三维目标与二维目标的最优匹配,输出高可靠融合结果。

1 理论基础

1.1 算法总体结构

基于激光点云与图像信息融合的交通环境车辆检测方法的总体结构如图1所示,主要包括以下几个步骤:

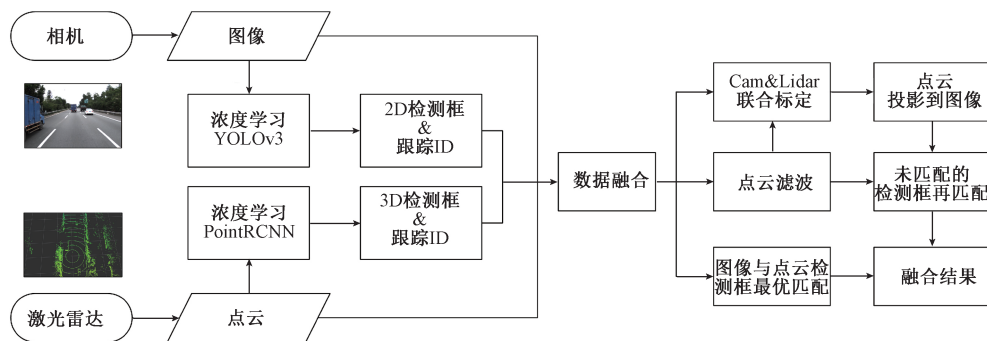


图1 融合算法总体架构

Fig.1 The architecture of the fusion algorithm

1) 利用深度学习算法分别对输入图像及激光点云进行目标检测。其中对图像采用 YOLOv3 算法获取检测框类别、中心点像素坐标及长宽尺寸信息;对点云采用 PointRCNN 算法获取 3D 检测框类别、中心点空间坐标及长宽高尺寸信息。

2) 利用匈牙利算法分别对前后帧的检测框及点云检测框做最优匹配,实现检测框的目标跟踪,分别建立图像及点云检测目标的跟踪序列。

3) 对图像及其检测框、点云及其检测框的跟踪序列进行融合,其中包括对点云滤波后投影到图像上、利用匈牙利算法对图像检测框及点云检测框做最优匹配,对未匹配的检测框做二次筛选等。

4) 根据融合后的检测框是否被匹配、被跟踪等特征,对所有目标检测框进行分类后输出融合结果,并筛选可靠性较高的检测框作为最终融合结果进行输出。

1.2 目标跟踪算法

融合算法利用匈牙利算法同时对图像与点云检测结果进行连续帧目标跟踪^[18-21],根据检测结果是否被跟踪来进一步提高检测精度。

目标跟踪算法架构如图 2 所示,包括以下步骤:

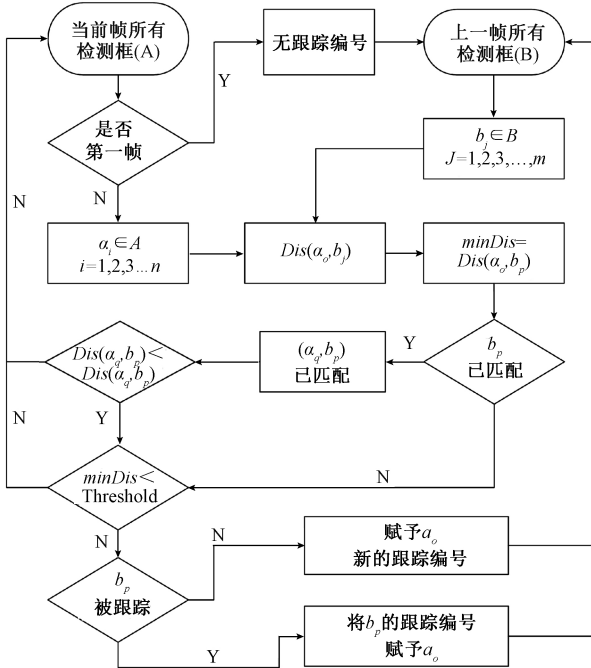


图 2 前后帧目标跟踪算法架构

Fig.2 The Architecture of the target tracking algorithm for continuous frames

1) 提取当前帧检测框 $a_i \in A (i = 1, 2, 3, \dots, n)$ 与上一帧的检测框 $b_j \in B (j = 1, 2, 3, \dots, m)$;

2) 计算当前帧检测框 $a_o \in A (o = 1, 2, 3, \dots, n)$ 与上一帧每一个检测框 b_j 的相对距离值 $Dis(a_o, b_j)$, 并提取

最小距离值 $\min Dis = Dis(a_o, b_p)$;

3) 若 b_p 为首次计算最小距离值,即 b_p 未匹配,则执行步骤 5);

4) 若 b_p 非首次计算最小距离值,即 (a_q, b_p) 已匹配,若 $Dis(a_o, b_p) < Dis(a_q, b_p)$, 则执行步骤 5), 否则进入步骤 2) 开始搜索下一个检测框。

5) 若最小距离值小于设定的阈值,即 $\min Dis < Threshold$, 且 b_p 在上一帧被跟踪,则将 b_p 的跟踪编号赋予 a_o , 否则赋予 a_o 新的跟踪编号。

通过上述步骤对前后帧检测框进行最优匹配,能够对连续检测到的目标实现目标跟踪,同时,对于偶尔出现的误检测框,由于无法匹配上一帧的检测结果,将取消其跟踪编号。因此,该方法能够在一定程度上排除误检测的目标,提高检测精度。

1.3 激光点云与图像信息融合算法

为了提高检测精度,降低误检率,需要利用目标检测算法分别对图像和点云进行目标检测,并输出检测结果进行目标跟踪及最优匹配。为了进一步提高图像和点云检测结果的匹配率,需要对点云及图像进行联合标定后将点云投影到图像上进行像素级融合,并将未匹配的图像检测框与投影到图像上的点云像素点进行再匹配,从而提高检测结果匹配率。

1) 激光雷达与相机联合标定

在 KITTI 数据集^[22]中,同时发布了所使用的 Velodyne-64 线激光雷达到不同相机之间的转换矩阵,包括激光雷达到 0 号灰度相机的外参矩阵、0 号相机的矫正旋转矩阵、其他相机的内参矩阵等。本文使用的检测数据包括 Velodyne-64 线激光雷达采集的点云数据及 2 号彩色相机采集的图像数据,因此,需要通过转换矩阵实现点云空间坐标到图像像素坐标的投影^[23-25],及图像像素坐标到点云空间坐标的反向推导。

激光雷达到相机的转换矩阵表示为 $T_{lidar}^{cam}(3 \times 4)$, 其由旋转矩阵 $R_{lidar}^{cam}(3 \times 3)$ 与平移矩阵 $T_{lidar}^{cam}(3 \times 1)$ 组成,0 号灰度相机的矫正旋转矩阵表示为 $R_{rect}^0(3 \times 3)$, 该矩阵近似为单位矩阵,2 号彩色相机的内参矩阵表示为 $P_{rect}^2(3 \times 4) \circ$

为了实现点云空间坐标与图像像素坐标的相互转换,需要各变换矩阵是可逆矩阵。因此,对各变换矩阵全部扩展成 4×4 的方阵,具体变化如下:

$$T_{lidar}^{cam} = \begin{pmatrix} R_{lidar}^{cam}(3 \times 3) & T_{lidar}^{cam}(3 \times 1) \\ 0 & 1 \end{pmatrix}_{4 \times 4} \quad (1)$$

$$R_{rect}^0 = \begin{pmatrix} R_{rect}^0(3 \times 3) & 0 \\ 0 & 1 \end{pmatrix}_{4 \times 4} \quad (2)$$

$$P_{rect}^2 = \begin{pmatrix} f_u & 0 & c_u & -f_u b_x \\ 0 & f_v & c_v & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}_{4 \times 4} \quad (3)$$

式中: f_u 和 f_v 分别表示相机在 u 轴和 v 轴方向上的尺度因子; C_u 和 C_v 分别表示图像主点在 u 轴和 v 轴方向上的偏移量; b_x 表示 2 号彩色相机与 0 号灰度相机的基线。

若点云空间坐标表示为 $\mathbf{X}$, 图像像素坐标表示为 $\mathbf{Y}$, 则点云空间坐标到图像像素坐标的投影计算方式如下:

$$\mathbf{Y} = \mathbf{P}_{rect}^2 \cdot \mathbf{R}_{rect}^0 \cdot \mathbf{Tr}_{lidar}^{cam} \cdot \mathbf{X} \quad (4)$$

式中: $\mathbf{X} = (x \ y \ z \ 1)^T$, $\mathbf{Y} = (Ku \ Kv \ K \ 1)^T$, K 为与空间 x 轴坐标相关的特定系数。激光雷达点云的空间三维坐标 (x, y, z) 经过上式的矩阵变换后即可求得对应的图像像素点坐标 (u, v) 。

本文融合算法在图像检测结果与点云检测结果匹配条件过程中, 同时考虑了图像检测框预估的深度信息与点云检测框测量的深度信息之间的关系。因此, 需要将图像检测框的像素坐标转换成激光雷达的空间坐标, 从而获取图像检测框预估的深度信息 d_{depth} 。由于激光雷达在车上位置固定, 即相对于水平地面的高度 h 固定, 因此选择图像检测框的底边中心点 (u_0, v_0) 作为检测目标下方的地面点 (x_0, y_0, z_0) , 即该像素点所对应的空间 z 轴方向坐标已知, $Z_0 = -h$ 。

根据激光雷达与基准相机的相对位置关系, $\mathbf{Tr}_{lidar}^{cam}$ 近似可表示为:

$$\mathbf{Tr}_{lidar}^{cam} = \begin{pmatrix} 0 & -1 & 0 & \mathbf{T}_x \\ 0 & 0 & -1 & \mathbf{T}_y \\ 1 & 0 & 0 & \mathbf{T}_z \\ 0 & 0 & 0 & 1 \end{pmatrix}_{4 \times 4} \quad (5)$$

由于 $\mathbf{R}_{rect}^0$ 近似为单位矩阵, 系数 K 可近似表示为 $K = x + \mathbf{T}_z$ 。

从图像像素点坐标到空间坐标点的转换公式可表示为:

$$\mathbf{X} = \mathbf{Tr}_{lidar}^{cam}{}^{-1} \cdot \mathbf{R}_{rect}^0{}^{-1} \cdot \mathbf{P}_{rect}^2{}^{-1} \cdot \mathbf{Y} \quad (6)$$

$$\mathbf{Tr}_{cam}^{lidar} = \mathbf{Tr}_{lidar}^{cam}{}^{-1} \cdot \mathbf{R}_{rect}^0{}^{-1} \cdot \mathbf{P}_{rect}^2{}^{-1} = \begin{pmatrix} T_{00} & T_{01} & T_{02} & T_{03} \\ T_{10} & T_{11} & T_{12} & T_{13} \\ T_{20} & T_{21} & T_{22} & T_{23} \\ 0 & 0 & 0 & 1 \end{pmatrix}_{4 \times 4} \quad (7)$$

$$x_0 = \frac{z_0 - T_{23}}{(T_{20}u_0 + T_{21}v_0 + T_{22})} - T_z \quad (8)$$

$$y_0 = (x_0 + \mathbf{T}_z)(T_{10}u_0 + T_{11}v_0 + T_{12}) + T_{13} \quad (9)$$

$$d_{depth} = \sqrt{x_0^2 + y_0^2} \quad (10)$$

通过上述坐标转换公式, 可以实现点云空间坐标与图像像素坐标的相互转换, 进而实现点云到图像的投影及图像检测框到三维世界坐标系的深度估计。

2) 融合方法

本文的多传感器融合算法在图像与点云的检测结果的基础上, 结合视觉感知与雷达感知各自的优势, 进一步

提高检测精度。算法流程如下:

(1) 将激光点云投影至图像

利用激光雷达与摄像头联合标定所得到的转换矩阵, 将点云数据投影到图像上, 可得到三维空间坐标 (x, y, z) 与二维像素坐标 (u, v) 的一一对应关系, 使得图像检测框内的目标具备激光雷达测量得到的空间坐标信息, 从而给图像检测框赋予更为准确的深度信息。

由于激光雷达每一帧所采集的数据量极其庞大, 为了提高检测算法计算速度, 提升实时性, 需要对点云数据进行降采样。本文算法对点云数据分别进行直通滤波与体素滤波, 仅保留摄像头视野范围内的点云数据, 并降低点云密度, 在满足测量精度的同时提高计算速度, 保证算法的实时性。

(2) 图像检测框与点云检测框的最优匹配

在每一帧的图像与点云都分别存在若干个检测框, 在这些检测框中, 存在多种检测情况, 包括正确检测且完成跟踪的目标、误检测目标等。因此, 对图像与点云的检测框利用匈牙利算法进行最优匹配, 实现检测结果的分类, 并择优输出, 能够在一定程度上提高检测结果准确性。匹配的具体步骤如下:

①提取图像检测框 $x_i \in \mathbf{X} (i = 1, 2, 3, \dots, n)$ 与点云检测框 $y_j \in \mathbf{Y} (j = 1, 2, 3, \dots, m)$;

②计算图像检测框 $x_u \in \mathbf{X} (u = 1, 2, 3, \dots, n)$ 与点云每一个检测框 y_j 的相对距离值 $Dis(x_u, y_j)$, 并提取最小距离值 $minDis = Dis(x_u, y_v)$;

③若 y_v 为首次计算最小距离值, 即 y_v 未匹配, 则执行步骤⑤;

④若 y_v 非首次计算最小距离值, 即 (x_u, y_v) 已匹配, 若 $Dis(x_u, y_v) < Dis(x_w, y_v)$, 则执行步骤⑤, 否则进入步骤②开始搜索下一个图像检测框;

⑤若最小距离值小于设定的阈值, 即 $minDis < Threshold$, 则将 (x_u, y_v) 进行匹配, 输入出融合结果 7, 并将 y_v 放入已匹配序列 Z 中保存。

此时, 该目标在图像和点云中同时被检测, 可作为准确度等级最高的融合结果进行输出。

(3) 未匹配的检测框二次筛选

如图 3 所示, 对于未匹配的图像检测框和点云检测框, 主要存在几种情况, 包括目标在点云中被检测到, 但在图像中受遮挡而未被检测; 目标在图像中被检测出, 但在点云中由于距离太远、点云数量较少而未被检测; 误检测目标等。

为提高目标检测率和检测精度, 对上一步未匹配的图像及检测框进行二次筛选, 具体步骤如下:

①提取为匹配的点云框 y_k , 若 y_k 被跟踪, 且 $y_k \in \mathbf{Z}$, 可认为 y_k 是因为在图像中受遮挡导致未匹配, 将 y_k 作为融合结果 1 输出。

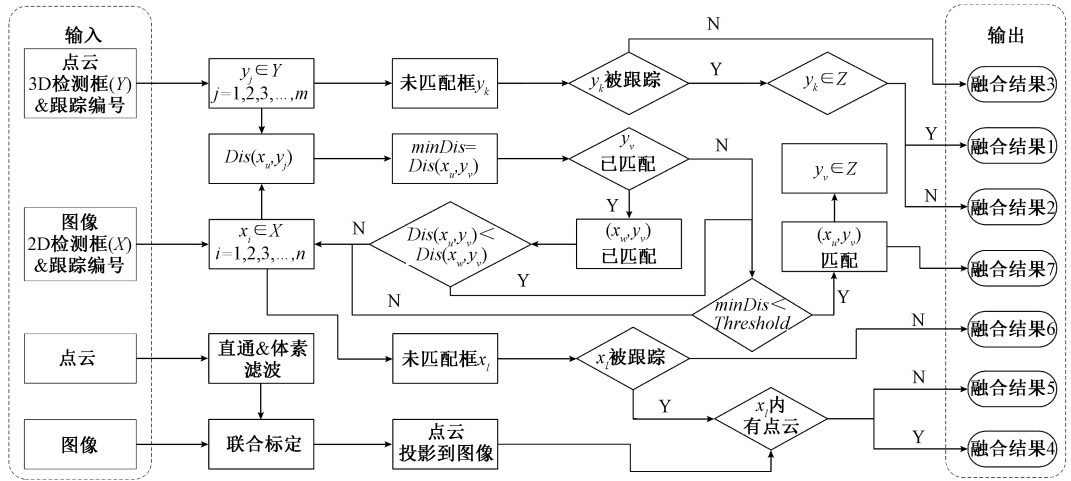


图3 图像与点云融合算法架构

Fig.3 The architecture of the image and point cloud fusion algorithm

②提取未匹配的图像框 x_i ，若 x_i 被跟踪且框内含有投影到图像的点云，可认为 x_i 是因为距离太远、点云数量较少而未被检测，将 x_i 作为融合结果4输出。

通过上述对未匹配的图像及点云检测框进行二次筛选，将筛选过后的图像及点云检测结果作为准确度等级次高的融合结果进行输出。

(4)对检测的结果进行分类并输出最终融合列表。

根据图3所示融合方法，可将图像与点云检测框分为7种类别进行输出，对融合结果进行分类输出后，筛选上述目标检测准确度等级最高及次高的结果作为最终融合结果，即同时在图像和点云中检测出来的目标、仅在跟踪序列中且匹配过的点云检测框、在跟踪序列且含有投影点云的图像检测框。将其余不在跟踪序列且无法相互匹配的检测框排除，能够大大降低目标误检率，提高检测精度。

2 试验验证

由于构建数据集工程较大且时间冗长，本文利用公开数据集 KITTI 及随机截取的实车测试结果作为依据来验证所提融合方法的性能。挑选 KITTI 跟踪数据集中带标签的第20个跟踪序列作为算法验证数据集。该序列数据集总共包括837帧公路采集的图像和点云数据，带标签的目标主要为公路上行驶的车辆，符合所提融合算法应用场景。

在 KITTI 跟踪数据集上的测试结果见表1，在837帧数据中包含了6845个标签，即数据真实值（ground truth），分别用图像目标检测算法 YOLOv3、点云目标检测算法 PointRCNN、及本文融合算法对该数据集每一帧数据进行测试，设置测试交并比阈值（intersection over

union, IoU）为0.5，即当所预测的检测框和真实检测框的交集和并集之间的比例大于0.5即 $\text{IoU} > 0.5$ 时，认为该检测框为正确的检测结果（true positives, TP），若 $\text{IoU} < 0.5$ ，则认为该检测框为错误的检测结果（false positives, FP）。统计正确检测结果及错误检测结果的数量即可计算精度（Precision）和召回率（Recall）及平均精度均值（mean average precision, mAP），计算公式如下：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (11)$$

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{TP}{\text{GroundTruth}} \quad (12)$$

$$\text{Precision}_c = \frac{N(TP)_c}{N(Total)_c} \quad (13)$$

$$\text{AP}_c = \frac{\sum \text{Precision}_c}{N(TotalImage)_c} \quad (14)$$

$$\text{mAP} = \frac{\sum \text{AP}_c}{N(Classes)} \quad (15)$$

式中： FN 表示正样本被错误分为负样本的个数（false negative, FN）， Precision_c 为图像中检测到的车辆数量 $N(TP)_c$ 与所有车辆数量 $N(Total)_c$ 之比， AP_c 为车辆的所有精度之和 $\sum \text{Precision}_c$ 与所有含有车辆的图像数量 $N(TotalImage)_c$ 之比， $N(Classes)$ 为类别数即数量为1。

可以看出，本文融合算法相比单独应用 YOLOv3 图像目标检测算法或 PointRCNN 点云目标检测算法，在检测精度上明显提升，误检率明显下降。融合算法相比 YOLOv3 算法精度提升16.43%，误检测目标数减少53.3%，总体 mAP 指标提升10.22%。相比 PointRCNN 算法精度提升12.01%，误检测目标数减少56.62%，总体

mAP 指标提升 3. 06%。

同时对比了两种不同的融合方法,包括对图像和点云的检测结果分别取交集或并集的融合方法,具体评价指标如表 1 所示。

表 1 算法性能指标对比 (IoU=0. 5)

Table 1 Comparison of algorithm performance indexes (IoU=0. 5)

方法	mAP/%	TP	FP	精度/%	召回率/%	ground truth
YOLOv3	51. 31	3 976	1 998	66. 56	58. 09	6 845
PointRCNN	58. 47	5 261	2 151	70. 98	76. 86	6 845
取交集融合	45. 27	3 395	286	92. 23	49. 60	6 845
取并集融合	67. 86	5 478	4 109	57. 14	80. 03	6 845
本文方法	61. 53	4 552	933	82. 99	66. 50	6 845

对检测结果取交集的方法,只考虑在图像和点云上共同检测到的目标,误检的目标数量在以上 5 种方法中最少,排除了大多数在图像或点云中的误检测目标,其检测精度最高,但与此同时,其融合的目标总数也最少,召回率及总体 mAP 指标在 5 种方法中最低。

对检测结果取并集的方法,同时考虑在图像和点云

上的所有检测结果,融合的目标总数最多,其检测的结果的召回率和总体 mAP 指标在以上 5 种方法中最高,但与此同时,该方法也囊括了来自图像和点云检测中的所有误检测目标,误检测的数量也大大提升,检测精度在 5 种方法中最低。

本文算法在对图像和点云检测结果取交集的基础上,根据图像和点云检测结果的跟踪信息,对准确度较高的融合结果进行择优输出,在保证检测精度的同时,大大降低误检率,相比其他方法而言更适用于无人驾驶汽车在交通环境下的车辆检测。

本文融合算法在机器人操作系统 (robot operating system, ROS) 上运行,输入点云与图像的帧率参照激光雷达采集频率为 10 Hz,并对输入的点云、图像、及 2D 检测框、3D 检测框等数据进行时间戳同步处理,平均单帧检测与融合时间 0. 1 s,满足自动驾驶场景下的实时性要求。

检测效果如图 4~7 所示,在 RGB 彩色图片中蓝色框代表真实目标框,绿色框代表正确检测的目标框,红色框则代表误检测的目标框。在点云鸟瞰图 (bird's-eye view) 中,红色框代表真实目标框,蓝色框代表点云检测框,绿色框代表融合后的检测框。若检测框与真实框重合,则会被真实框覆盖。



图 4 点云到图像的投影
Fig.4 Projection from point cloud to image



图 5 YOLOv3 算法检测结果
Fig.5 The detection results of YOLOv3

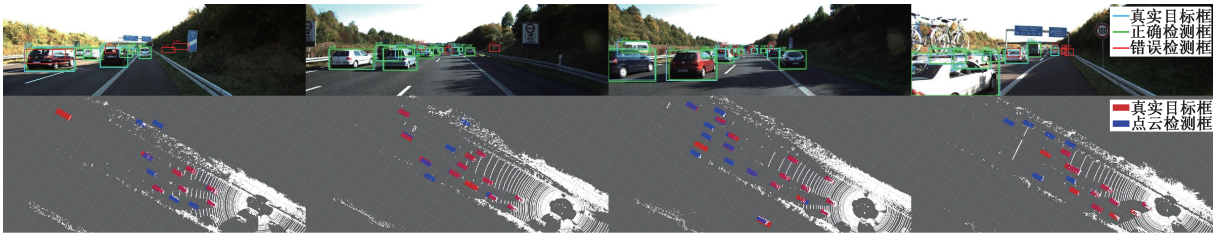


图 6 PointRCNN 算法检测结果
Fig.6 The detection results of PointRCNN

从以上检测效果图可以看出,图像检测结果与点云检测结果都存在不同程度的误检测,图像的误检更多发生在车辆被遮挡的情况下,检测框与真实框交并比小于

0. 5,导致误检。点云的误检更多是因为将道路两旁的灌木、树叶等与车辆点云形状相似的物体检测成车辆,造成误检。融合算法图像目标检测结果及点云目标检测结果

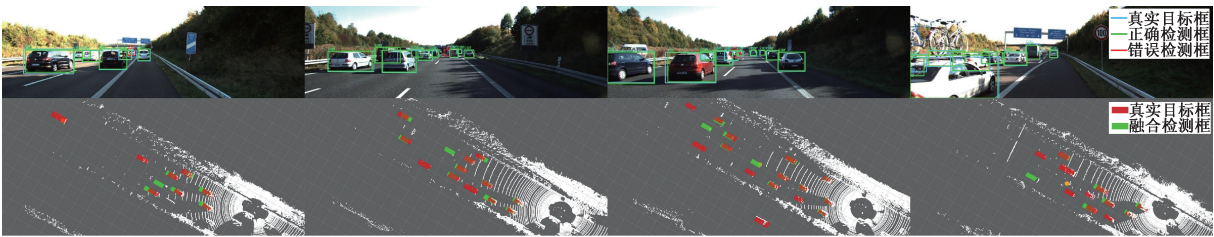


图 7 融合算法检测结果

Fig.7 The detection results of the fusion algorithm

进行匹配、融合并分类后输出,能够显著降低目标误检率,并在一定程度上提高检测精度。

为进一步验证所提融合方法的鲁棒性,利用如图 8 所示搭载激光雷达及相机的试验车,在广州某地区实际道路进行实时车辆检测实验,随机截取实车测试结果如图 9 所示。



图 8 试验车

Fig.8 The testing car

在图 9(b)中,图像检测结果受到光照影响而出现漏检测的情况,在图 9(c)中,点云检测结果受形状相似的干扰物影响而出现误检测现象,在图 9(d)中,经过本文融合方法处理后的检测结果能够有效避免当前环境下的漏检及误检现象。根据实验测试结果,在实际交通环境下所提融合方法车辆检测性能显著。

3 结 论

本文提出了一种基于激光点云与图像信息融合的交通环境车辆检测算法,通过对图像和点云原始数据分别进行目标检测与跟踪,利用匈牙利算法对检测结果进行最优匹配,并结合点云到图像的投影对未匹配的图像检测框进行进一步筛选,最终对融合列表进行分类后择优输出。通过 KITTI 公开数据集的部分连续交通场景和实车测试验证了所提融合方法的检测性能,实验结果表明,在满足实时性的前提下,本文方法检测结果相比未融合的结果精度指标提升 12% 以上、误检测目标数减少 50% 以上,总体 mAP 指标提升 3% 以上,能够为无人驾驶系统提供更可靠的环境感知信息,提升目标检测精度并减少误检率。

参考文献

[1] 张艳辉,徐坤,郑春花,等.智能电动汽车信息感知技术研究进展[J].仪器仪表学报,2017,38(4):794-805.
ZHANG Y H, XU K, ZHENG CH H, et al. Advanced research on information perception technologies of intelligent electric vehicles[J]. Chinese Journal of Scientific Instrument, 2017, 38(4): 794-805.

[2] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014:580-587.

[3] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

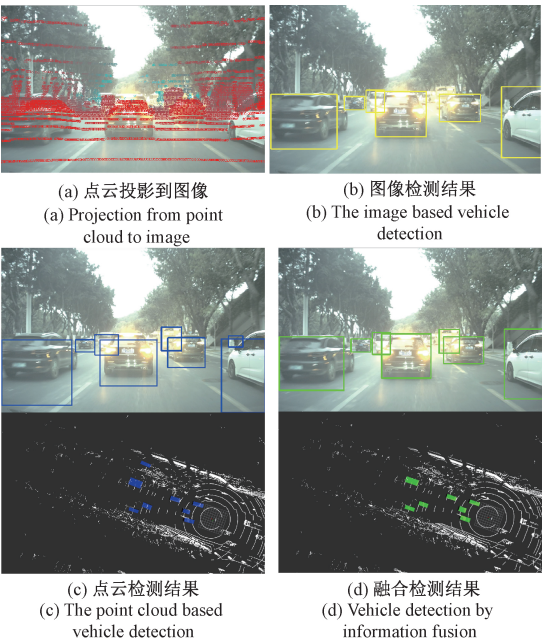


图 9 实际交通环境下的车辆检测结果

Fig.9 The vehicle detection in real traffic environment

- [4] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]. Advances in Neural Information Processing Systems, 2015: 91-99.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, Springer, 2016: 21-37.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [7] REDMON J, FARHADI A. Yolov3: An incremental improvement [J]. arXiv preprint arXiv: 1804.02767, 2018.
- [8] SHI S, WANG Z, WANG X, et al. Part-A² net: 3D part-aware and aggregation neural network for object detection from point cloud [J]. arXiv preprint arXiv: 1907.03670, 2019.
- [9] ZHOU Y, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3d object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4490-4499.
- [10] LANG A H, VORA S, CAESAR H, et al. PointPillars: Fast encoders for object detection from point clouds[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 12697-12705.
- [11] SHI S, WANG X, LI H. Pointrenn: 3d object proposal generation and detection from point cloud [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 770-779.
- [12] LIANG M, YANG B, CHEN Y, et al. Multi-task multi-sensor fusion for 3D object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 7345-7353.
- [13] LIANG M, YANG B, WANG S, et al. Deep continuous fusion for multi-sensor 3d object detection [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 641-656.
- [14] CHEN X, MA H, WAN J, et al. Multi-view 3D object detection network for autonomous driving [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1907-1915.
- [15] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2018: 1-8.
- [16] WEI P, CAGLE L, REZA T, et al. LiDAR and camera detection fusion in a real-time industrial multi-sensor collision avoidance system [J]. Electronics, 2018, 7(6): 84.
- [17] WU T EN, TSAI C C, GUO J I. Lidar/camera sensor fusion technology for pedestrian detection[C]. 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, 2017: 1675-1678.
- [18] 汪济洲, 鲁昌华, 蒋薇薇. 一种基于随机场多运动目标跟踪算法[J]. 电子测量与仪器学报, 2017, 31(6): 909-913.
WANG J ZH, LU CH H, JIANG W W. New multi-target tracking algorithm based on conditional random field[J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(6): 909-913.
- [19] TIAN J L, WANG Z Y, LI L, et al. Tracking human poses with head orientation estimation [J]. Instrumentation, 2017, 4(3): 40-46.
- [20] 余铎, 王耀南, 毛建旭, 等. 基于视觉的移动机器人目标跟踪方法[J]. 仪器仪表学报, 2019, 40(1): 227-235.
YU D, WANG Y N, MAO J X, et al. Vision-based object tracking method of mobile robot [J]. Chinese Journal of Scientific Instrument, 2019, 40(1): 227-235.
- [21] 孙宇嘉, 于纪言, 王晓鸣. 适用于复杂场景的多目标跟踪算法[J]. 仪器仪表学报, 2019, 40(3): 126-137.
SUN Y J, YU J Y, WANG X M. Multiple object tracking algorithm for the complex scenario[J]. Chinese Journal of Scientific Instrument, 2019, 40(3): 126-137.
- [22] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: The KITTI dataset [J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.
- [23] 夏鹏飞, 尹慧琳, 何艳侠. 基于最大互信息的激光雷达与相机的配准[J]. 仪器仪表学报, 2018, 39(1): 34-41.
XIA P F, YIN H L, HE Y X. Calibration of lidar and camera based on maximum mutual information [J]. Chinese Journal of Scientific Instrument, 2018, 39(1): 34-41.
- [24] 董方新, 蔡军, 解杨敏. 立体视觉和三维激光系统的联合标定方法[J]. 仪器仪表学报, 2017, 38(10): 2589-2596.
DONG F X, CAI J, XIE Y M. Joint calibration method for stereo vision system and 3D laser system[J]. Chinese Journal of Scientific Instrument, 2017, 38(10): 2589-2596.

- [25] 彭梦,蔡自兴.基于系数矩阵二范数的 2 维激光-摄标定方法[J].仪器仪表学报,2015,36(4):749-757.

PENG M, CAI Z X. Calibration method for 2D laser radar and camera based on 2-norm of coefficient matrix [J]. Chinese Journal of Scientific Instrument, 2015, 36(4): 749-757.

作者简介



郑少武,2018 年于华南理工大学获得学士学位,现为华南理工大学硕士研究生,主要研究方向为智能驾驶环境感知、多传感器融合。

E-mail: zhengswu@qq.com

Zheng Shaowu received his B. Sc. degree from South China University of Technology in 2018. He is currently a M. Sc. candidate at South China University of Technology. His main research interests include intelligent driving, environment perception and multi-sensor fusion.



李巍华(通信作者),分别在 1995 年和 1998 年于太原理工大学获得学士学位和硕士学位,2003 年于华中科技大学获得博士学位,现为华南理工大学教授,主要研究方向为智能驾驶、装备智能运维与健康管理、以及汽车 NVH 性能测试等。

E-mail: whlee@scut.edu.cn

Li Weihua (Corresponding author) received his B. Sc. and M. Sc. degree both from Taiyuan University of Technology in 1995 and 1998, respectively, and received his Ph. D. degree from Huazhong University of Science and Technology in 2003. He is currently a professor at South China University of Technology. His main research interests include intelligent driving, machinery intelligent maintenance and health management, vehicle NVH performance test.



胡坚耀,分别在 2010 年和 2012 年于武汉大学获得硕士学位,2015 年于中国科学院大学获得博士学位。现为工业和信息化部电子第五研究所新能源智能汽车团队负责人。主要研究方向为新能源智能汽车测试评价技术。

E-mail: hujianyao@ceppei.biz

Hu Jianyao received his B. Sc. and M. Sc. both from Wuhan University in 2010 and 2012, respectively, and received his Ph. D. degree in 2015 from University of Chinese Academy of Sciences. He is currently the leader of the new energy intelligent vehicle team at China CEPREI Laboratory. His main research interests include new energy intelligent vehicle test and evaluation technology.